

[illegible][illegible]

## **FIGURE 2**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA44686

><subunit 1 of 1, 246 aa, 1 stop

><MW: 26994, pI: 6.43, NX(S/T): 0

MLWRQLIYWQLLALFFLPFCLCQDEYMESPQTGGLPPDCSKCCHGDYSFRGYQGPPGPPGPP  
GIPGNHGNNGNNGATGHEGAKGEKGDKGDLGPRGERGQHGPKEKGYPGIPPELQIAFMASL  
ATHFSNQNSGIIFFSSVETNIGNFFDVMTGFRFGAPVSGVYFFTFSMMKHEDVEEVYVYLMHNG  
NTVFSMYSYEMKGKSDTSSNHAVLKLAKGDEVWLRMGNGALHGDHQRFFSTFAGFLLFETK

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-22

#### **Motif name: Clq domain signature.**

amino acids 137-167

#### **Clq domain proteins.**

amino acids 135-169, 202-221, 235-244, 57-91, 60-94, 54-88, 81-114, 78-111, 63-96, 51-84, 45-78, 48-81, 33-66, 66-99 and 42-75

Figure 2

### FIGURE 3

GAGAGAATAGCTACAGATTCTCCATCCTCAGTCTTTGCAAGGCGACAGCTGTGCCAGCCGGG  
CTCTGGCAGGCTCCTGGCAGCATGGAGTGAAGCTTGGGACCCTCCTGCTGGCCCTTGCCCT  
GGGCCTGGCCCAGCCAGCCTCTGCCCCGCCGGAAGCTGCTGGTGTTTCTGCTGGATGGTTTTTC  
GCTCAGACTACATCAGTGATGAGGCGCTGGAGTCATTGCCTGGTTTCAAAGAGATTGTGAGC  
AGGGGAGTAAAAGTGGATTACTTGACTCCAGACTTCCCTAGTCTCTCGTATCCCAATTATTA  
TACCCTAATGACTGGCCGCCATTGTGAAGTCCATCAGATGATCGGGAACACATGTGGGACC  
CCACCACCAACAAGTCCTTTGACATTGGCGTCAACAAAGACAGCCTAATGCCTCTCTGGTG  
AATGGATCAGAACCTCTGTGGGTCACTCTGACCAAGGCCAAAAGGAAGGTCTACATGTACTA  
CTGGCCAGGCTGTGAGGTTGAGATTCTGGGTGTCAGACCCACCTACTGCCTAGAATATAAAA  
ATGTCCCAACGGATATCAATTTTGCCAATGCAGTCAGCGATGCTCTTGACTCCTTCAAGAGT  
GGCCGGGCCGACCTGGCAGCCATATACCATGAGCGCATTGACGTGGAAGGCCACCACTACGG  
GCCTGCATCTCCGCAGAGGAAAGATGCCCTCAAGGCTGTAGACACTGTCCTGAAGTACATGA  
CCAAGTGGATCCAGGAGCGGGGCCTGCAGGACCGCCTGAACGTCATTATTTTCTCGGATCAC  
GGAATGACCGACATTTTCTGGATGGACAAAGTGATTGAGCTGAATAAGTACATCAGCCTGAA  
TGACCTGCAGCAAGTGAAGGACCGCGGGCCTGTTGTGAGCCTTTGGCCGGCCCCCTGGGAAAC  
ACTCTGAGATATATAACAACTGAGCACAGTGGAACACATGACTGTCTACGAGAAAGAAGCC  
ATCCCAAGCAGGTTCTATTACAAGAAAGGAAAGTTTGTCTCTCCTTTGACTTTAGTGGCTGA  
TGAAGGCTGGTTCATAACTGAGAATCGAGAGATGCTTCCGTTTTGGATGAACAGCACCGGCA  
GGCGGGAAGGTTGGCAGCGTGGATGGCACGGCTACGACAACGAGCTCATGGACATGCGGGGC  
ATCTTCCTGGCCTTCGGACCTGATTTCAAATCCAACCTCAGAGCTGCTCCTATCAGGTCGGT  
GGACGTCTACAATGTCATGTGCAATGTGGTGGGCATCACCCCGCTGCCCAACAACGGATCCT  
GGTCCAGGGTGATGTGCATGCTGAAGGGCCGCGCCGGCACTGCCCCGCCTGTCTGGCCCAGC  
CACTGTGCCCTGGCACTGATTCTTCTCTCCTGCTTGCATAACTGATCATATTGCTTGTCTC  
AGAAAAAACACCATCAGCAAAGTGGGCCTCCAAAGCCAGATGATTTTCATTTTATGTGTGA  
ATAATAGCTTCATTAACACAATCAAGACCATGCACATTGTAAATACATTATTCTTGGATAAT  
TCTATACATAAAAGTTCCTACTTGTTAAA

### Important features of the protein:

amino acids 1-22

amino acids 100-104, 118-122, 341-345, 404-408

amino acids 148-154, 365-371

## amino acids 343-347

[illegible]

## **FIGURE 5**

GCCAGGTGTGCAGGCCGCTCCAAGCCCAGCCTGCCCCGCTGCCGCCACC**ATG**ACGCTCCTCC  
CCGGCCTCCTGTTTCTGACCTGGCTGCACACATGCCTGGCCCACCATGACCCCTCCCTCAGG  
GGGCACCCCCACAGTCACGGTACCCACACTGCTACTCGGCTGAGGAACTGCCCCTCGGCCA  
GGCCCCCCCCACACCTGCTGGCTCGAGGTGCCAAGTGGGGGCAGGCTTTGCCTGTAGCCCTGG  
TGTCCAGCCTGGAGGCAGCAAGCCACAGGGGGAGGCACGAGAGGCCCTCAGCTACGACCCAG  
TGCCCGGTGCTGCGGCCGGAGGAGGTGTTGGAGGCAGACACCCACCAGCGCTCCATCTCACC  
CTGGAGATAACCGTGTGGACACGGATGAGGACCGCTATCCACAGAAGCTGGCCTTCGCCGAGT  
GCCTGTGCAGAGGCTGTATCGATGCACGGACGGGCCGCGAGACAGCTGCGCTCAACTCCGTG  
CGGCTGCTCCAGAGCCTGCTGGTGCTGCGCCGCCGGCCCTGCTCCCGCGACGGCTCGGGGCT  
CCCCACACCTGGGGCCTTTGCCTTCCACACCGAGTTCATCCACGTCCCCGTCGGCTGCACCT  
GCGTGCTGCCCCGTTCAGTG**TGA**ACCGCCGAGGCCGTGGGGCCCCCTAGACTGGACACGTGTGC  
TCCCCAGAGGGCACCCCCTATTTATGTGTATTTATTGTTATTTATATGCCTCCCCCAACACT  
ACCCTTGGGGTCTGGGCATTCCCCGTGTCTGGAGGACAGCCCCCACTGTTCTCCTCATCTC  
CAGCCTCAGTAGTTGGGGGTAGAAGGAGCTCAGCACCTCTTCCAGCCCTTAAAGCTGCAGAA  
AAGGTGTCACACGGCTGCCTGTACCTTGGCTCCCTGTCCTGCTCCCGGCTTCCCTTACCCTA  
TCACTGGCCTCAGGCCCCGCAGGCTGCCTCTTCCCAACCTCCTTGGAAGTACCCCTGTTTCT  
TAAACAATTATTTAAGTGTACGTGTATTATTAACTGATGAACACATCCCCAAAA

[illegible]

### Important features:

amino acids 1-18

amino acids 112-121

amino acids 32-38, 55-61, 133-139

amino acids 3-25

amino. acids 99-195

## **FIGURE 7**

CGGCCAGGGCGCCGACAGCCCGACCTCACCAGGAGAAC**ATG**CAGCTCGGCACTGGGCTCCTG  
CTGGCCGCGCTCCTGAGCCTGCAGCTGGCTGCAGCCGAAGCCATATGGTGTCAACAGTGCAC  
GGGCTTCGGAGGGTGCTCCCATGGATCCAGATGCCTGAGGGACTCCACCCACTGTGTACCA  
CTGCCACCCGGGTCCTCAGCAACACCGAGGATTTGCCTCTGGTCACCAAGATGTGCCACATA  
GGCTGCCCCGATATCCCCAGCCTGGGCCTGGGCCCCCTACGTATCCATCGCTTGCTGCCAGAC  
CAGCCTCTGCAACCATGAC**TGA**CGGCTGCCCTCCTCCAGGCCCCCGGACGCTCAGCCCCAC  
AGCCCCACAGCCTGGCGCCAGGGCTCACGGCCGCCCCCTCCCTCGAGACTGGCCAGCCCACC  
TCTCCCGGCCTCTGCAGCCACCGTCCAGCACCGCTTGTCTAGGGAAGTCCTGCGTGGAGTC  
TTGCCTCAATCTGCTGCCGTCCAAGCCTGGGGCCCATCGTGCCTGCCGCCCCCTCAGGTCCC  
GACCTCCCCACAATAAAATGTGATTGGATCGTGTGGTACAAAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Figure 7

## **FIGURE 8**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA77623

><subunit 1 of 1, 97 aa, 1 stop

><MW: 10160, pI: 6.56, NX(S/T): 0

MQLGTGLLLAAVLSLQLAAAEAIWCHQCTGFGGCSHGSRCLRDSTHCVTTATRVLSNTEDLP  
LVTKMCHIGCPDIPSLGLGPYVSIACCQTSLCNHD

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-20

#### **N-myristoylation sites.**

amino acids 6-11 and 33-38

#### **Prokaryotic membrane lipoprotein lipid attachment sites.**

amino acids 24-34 and 78-88

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100



## FIGURE 9

CCAGGACCAGGGCGCACCGGCTCAGCCTCTCACTTGTGTCAGAGGCCGGGGAAGAGAAGCAAAG  
CGCAACGGTGTGGTCCAAGCCGGGGCTTCTGCTTCGCCTCTAGGACATACACGGGACCCCT  
AACTTCAGTCCCCCAAACGCGCACCCCTCGAAGTCTTGAAGTCCAGCCCCGCACATCCACGCG  
CGGCACAGGCGCGGCAGGCGGCAGGTCCCGGCCGAAGGCGATGCGCGCAGGGGGTTCGGGCAG  
CTGGGCTCGGGCGGCAGGAGTAGGGCCCCGCGCAGGGAGGCAGGGAGGCTGCATATTCAGAGTC  
GCGGGCTGCGCCCTGGGCAGAGGCCGCCCTCGCTCCACGCAACACCTGCTGCTGCCACCGCG  
CCGCGATGAGCCGCGTGGTCTCGCTGCTGCTGGGCGCCGCGCTGCTCTGCGGCCACGGAGCC  
TTCTGCCGCGCGTGGTCAAGCGGCCAAAAGGTGTGTTTTGCTGACTTCAAGCATCCCTGCTA  
CAAAATGGCCTACTTCCATGAACTGTCCAGCCGAGTGAGCTTTCAGGAGGCACGCCTGGCTT  
GTGAGAGTGAGGGAGGAGTCCCTCCTCAGCCTTGAGAATGAAGCAGAACAGAAGTTAATAGAG  
AGCATGTTGCAAAACCTGACAAAACCCGGGACAGGGATTCTGATGGTGATTTCTGGATAGG  
GCTTTGGAGGAATGGAGATGGGCAAACATCTGGTGCCTGCCCAGATCTCTACCAAGTGGTCTG  
ATGGAAGCAATTCAGTACCGAACTGGTACACAGATGAACCTTCCTGCGGAAGTGAAAAG  
TGTGTTGTGATGTATCACCAACCAACTGCCAATCCTGGCCTTGGGGGTCCCTACCTTTACCA  
GTGGAATGATGACAGGTGTAACATGAAGCACAATTATATTTGCAAGTATGAACCAGAGATTA  
ATCCAACAGCCCCCTGTAGAAAAGCCTTATCTTACAAATCAACCAGGAGACACCCATCAGAAT  
GTGGTTGTTACTGAAGCAGGTATAATTCCCAATCTAATTTATGTTGTTATACCAACAATACC  
CCTGCTCTTACTGATACTGGTTGCTTTTGGAACTGTTGTTTCCAGATGCTGCATAAAAAGTA  
AAGGAAGAACAAAACTAGTCCAAACCAGTCTACACTGTGGATTTCAAAGAGTACCAGAAAA  
GAAAGTGGCATGGAAGTATAAATAACTCATTGACTTGGTTCAGAAATTTTGTAAATCTGGATC  
TGTATAAGGAATGGCATCAGAACAAATAGCTTGGAAATGGCTTGAAATCACAAAGGATCTGCAA  
GATGAACTGTAAGCTCCCCCTTGAGGCAAATATTAAAGTAATTTTTATATGTCTATTATTTT  
ATTTAAAGAATATGCTGTGCTAATAATGGAGTGAGACATGCTTATTTTGCTAAAGGATGCAC  
CCAACTTCAAACCTTCAAGCAAATGAAATGGACAATGCAGATAAAGTTGTTATCAACACGTC  
GGGAGTATGTGTGTTAGAAGCAATTCCTTTTATTTCTTTACCTTTTACATAAGTTGTTATCTA  
GTCAATGTAATGTATATTGTATTGAAATTTACAGTGTGCAAAAGTATTTTACCTTTGCATAA  
GTGTTTGATAAAAATGAACTGTTCTAATATTTATTTTATGGCATCTCATTTTCAATACAT  
GCTCTTTTGATTAAAGAACTTATTACTGTTGTCAACTGAATTCACACACACACAAATATAG  
TACCATAGAAAAAGTTTGTTTTCTCGAAATAATTCATCTTTCAGCTTCTCTGCTTTTGGTCA  
ATGTCTAGGAAATCTCTCAGAAATAAGAAGCTATTTTCAATTAAGTGTGATATAAACCTCCTC  
AAACATTTTACTTAGAGGCAAGGATTGTCTAATTTCAATTTGTGCAAGACATGTGCCTTATAA  
TTATTTTGTAGCTTAAATTAACAGATTTTGTAAATAATGTAACCTTTGTTAATAGGTGCATAA  
ACACTAATGCAGTCAATTTGAACAAAAGAAGTGACATACACAATATAAATCATATGTCTTCA  
CACGTTGCCTATATAATGAGAAGCAGCTCTCTGAGGGTTCTGAAATCAATGTGGTCCCTCTC  
TTGCCCACTAAACAAAGATGGTTGTTTCGGGGTTTGGGATTGACACTGGAGGCAGATAGTTGC  
AAAGTTAGTCTAAGGTTTCCCTAGCTGTATTTAGCCTCTGACTATATTAGTATACAAAGAGG  
TCATGTGGTTGAGACCAGGTGAATAGTCACTATCAGTGTGGAGACAAGCACAGCACACAGAC  
ATTTTAGGAAGGAAAGGAACACGAAATCGTGTGAAAATGGGTTGGAACCCATCAGTGATCG  
CATATTCATTGATGAGGGTTTGGCTTGAGATAGAAAATGGTGGCTCCTTTCTGTCTTATCTCC  
TAGTTTCTTCAATGCTTACGCCTTGTTCTTCTCAAGAGAAAGTTGTAACCTCTCTGGTCTTCA  
TATGTCCCTGTGCTCCTTTTAACCAATAAAGAGTTCTTGTCTTCTGGGGGAAAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAA

## **FIGURE 10**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA79230

><subunit 1 of 1, 273 aa, 1 stop

><MW: 30431, pI: 6.79, NX(S/T): 3

MSRVVSLLLGAALLCGHGAFCRRVVSQGKVCFADFHKPCYKMAYFHELSSRVSFQEARLACE  
SEGGVLLSLENEAEQKLIESMLQNLTKPGTGISDGDFWIGLWRNGDGQTSGACPDLYQWSDG  
SNSQYRNWYTDEPSCGSEKCVVMYHQPTANPGLGGPYLYQWNDDRCNMKHNYICKYEPEINP  
TAPVEKPYLTNQPGDTHQNVVVTEAGIIPNLIYVVIPTIPLLLLILVAFGTCCFQMLHKS  
RTKTSPNQSTLWISKSTRKESGMEV

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-21

#### **Transmembrane domain:**

amino acids 214-235

#### **N-glycosylation sites.**

amino acids 86-89 and 255-258

#### **cAMP- and cGMP-dependent protein kinase phosphorylation site.**

amino acids 266-269

#### **N-myristoylation sites.**

amino acids 27-32, 66-71, 91-96, 93-98, 102-107, 109-114, 140-145  
and 212-217

[illegible]

## **FIGURE 12**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA79862

><subunit 1 of 1, 571 aa, 1 stop

><MW: 62282, pI: 5.56, NX(S/T): 1

MERAVRVESGVLVGVVCLLLACPATATGPEVAQPEVDTTLGRVRGRQVGVGKTDRLVNVFLG  
IPFAQPPLGPDREFSAPHPAQPWEGVRDASTAPPMCLQDVESMNSSRFVLNGKQQIFSVSEDC  
LVLNVYSPA EVPAGSGRPVMVWVHGGALITGAATSYDGSALAA YGDVVVVTVQYRLGVLGFF  
STGDEHAPGNQGF LDVVAALRWVQENIAPFGGDLNCVTVFGGSAGGSIIISGLVLSPVAAGLF  
HRAITQSGVITTPGIIDSHPWPLAQKIAN TLACSSSSPAEMVQCLQQKEGEELVLSKKLKNT  
IYPLTVDGT VFPKSPKELLKEKPFHSPFLMGVNNHEFSWLI PRGWGLLD TMEQMSREDMLA  
ISTPVLTS LDVPPPEMMPTVIDEYLG SNSDAQAKCQAFQEFMGDVFINVPTVSFSRYLRDSGS  
PVFFYEFQHRPSSFAKIKPAWVKADHGAEGAFVFGGPFLMDESSRLAFPEATEEEKQLSLTM  
MAQWTHFARTGDPNSKALPPWPQFNQAEQYLEINPVPRAGQKFREAWMQFWSETLPSKIQQW  
HQKQKNRKAQEDL

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-27

#### **Transmembrane domain:**

amino acids 226-245

#### **N-glycosylation site.**

amino acids 105-109

#### **N-myristoylation sites.**

amino acids 10-16, 49-55, 62-68, 86-92, 150-156, 155-161,  
162-168, 217-223, 227-233, 228-234, 232-238, 262-268, 357-363,  
461-467

#### **Prokaryotic membrane lipoprotein lipid attachment site.**

amino acids 12-23

#### **Carboxylesterases type-B serine active site.**

amino acids 216-232

## FIGURE 13

CATGGAGCCTCTTGCAGCTTACCCGCTAAAATGTTCCGGGCCCAGAGCAAAGGTATTTGCAG  
TTTTGCTGTCTATAGTTCTATGCACAGTAACGCTATTTCTTCTACAATAAAATTCCTCAA  
CCTAAAATCAACAGCTTTTATGCCTTTGAAGTGAAGGATGCAAAAGGAAGAACTGTTTCTCT  
GGAAAAGTATAAAGGCAAAGTTTCACTAGTTGTAAACGTGGCCAGTGACTGCCAACTCACAG  
ACAGAAATTACTTAGGGCTGAAGGAACTGCACAAAGAGTTTGGACCATCCCACTTCAGCGTG  
TTGGCTTTTCCCTGCAATCAGTTTGGAGAATCGGAGCCCCGCCCAAGCAAGGAAGTAGAATC  
TTTTGCAAGAAAAAACTACGGAGTAACTTTCCCCATCTTCCACAAGATTAAGATTCTAGGAT  
CTGAAGGAGAACCTGCATTTAGATTTCTTGTGATTCTTCAAAGAAGGAACCAAGGTGGAAT  
TTTTGGAAGTATCTTGTCAACCCTGAGGGTCAAGTTGTGAAGTTCTGGAGGCCAGAGGAGCC  
CATTGAAGTCATCAGGCCTGACATAGCAGCTCTGGTTAGACAAGTGATCATAAAAAAGAAAG  
AGGATCTATGAGAATGCCATTGCGTTTCTAATAGAACAGAGAAATGTCTCCATGAGGGTTTG  
GTCTCATTTTAAACATTTTTTTTTTTTGGAGACAGTGTCTCACTCTGTCACCCAGGCTGGAGTG  
CAGTAGTGCGTTCTCAGCTCATTGCAACCTCTGCCTTTTTTAAACATGCTATTAAATGTGGCA  
ATGAAGGATTTTTTTTTTAATGTTATCTTGCTATTAAGTGGTAATGAATGTTCCCAGGATGAG  
GATGTTACCCAAAGCAAAAATCAAGAGTAGCCAAAGAATCAACATGAAATATATTAATACTACT  
TCCTCTGACCATACTAAAGAATTCAGAATACACAGTGACCAATGTGCCTCAATATCTTATTG  
TTCAACTTGACATTTTCTAGGACTGTACTTGATGAAAATGCCAACACACTAGACCACTCTTT  
GGATTCAAGAGCACTGTGTATGACTGAAATTTCTGGAATAACTGTAAATGGTTATGTTAATG  
GAATAAAACACAAATGTTGAAAAATGTAAAATATATATACATAGATTCAAATCCTTATATAT  
GTATGCTTGTTTTGTGTACAGGATTTTGTTTTTTCTTTTTTAAGTACAGGTTCCTAGTGTTTT  
ACTATAACTGTCACTATGTATGTAAGTACATATATAAATAGTCATTTATAAATGACCGTAT  
TATAACATTTGAAAAAGTCTTCATCAAAAAA

## **FIGURE 14**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA80136

><subunit 1 of 1, 209 aa, 1 stop

><MW: 23909, pI: 9.68, NX(S/T): 0

MEPLAAYPLKCSGPRAKVFAVLLSIVLCTVTFLFLQLKFLKPKINSFYAFEVKDAKGRTVSL  
EKYKGKVSLLVNVASDCQLTDRNYLGLKELHKEFGPSHFVLAFCNQFGESEPRPSKEVES  
FARKNYGVTFPIFHKIKILGSEGEPAFRFLVDSSKKEPRWNFWKYLVNPEGQVVKFWRPEEP  
IEVIRPDIAALVRQVIKKKEDL

**Important features of the protein:**

**Signal peptide:**

amino acids 1-31

**Glutathione peroxidases signature 2.**

amino acids 104-112

**Glutathione peroxidases.**

amino acids 57-82

[illegible]

TGTCGCGCTGGCCCTCGCC**ATG**CAGACACCCCGCAGAGCGTCCCCCTCCCCGCCCGGCCCTCCTGCTTCTGCTGCTGCTA  
CTGGGGGGGCGCCCACGGCCTCTTTCTGAGGAGCCGCCGCCGCTTAGCGTGGCCCCCAGGGACTACCTGAACCAC  
TATCCCGTGTTTGTGGGCAGCGGGCCCGACGCCTGACCCCGCAGAAGGTGCTGACGACCTCAACATCCAGCGA  
GTCCTGCGGGTCAACAGGACGCTGTTTCATTGGGGACAGGGACAACCTCTACCGCGTAGAGCTGGAGCCCCCAGC  
TCCACGGAGCTGCGGTACACAGAGAACTGACCTGGAGATCTAACCCAGCACATAAACGTGTGTCCGATGAAG  
GGCAAAACAGGAGGGCGAGTGTGAAACTTCGTAAGGTGCTGCTCCTTCGGGACGAGTCCACGCTCTTTGTGTGC  
GGTTCCAACGCCTTCAACCCGGTGTGCGCCAACCTACAGCATAGACACCCTGCAGCCCGTTCGGAGACAACATCAGC  
GGTATGGCCCGCTGCCCCGTACGACCCCCAAGCACGC~~CA~~ATGTTGCCCTCTTCTCTGACGGGATGCTCTTCACAGCT  
ACTGTTACCGACTTCTTAGCCATTGATGCTGTCACTTACCGCAGCCTCGGGGACAGGCCACCCTGCGCACCGT  
AAACATGACTCAAAGTGGTTCAAAGAGCCTTACTTTTGTCATGCGGTGGAGTGGGGACGCCATGTCTACTTCTTC  
TTCGGGAGATTTCGATGGAGTTTAACTACCTGGAGAAGGTGGTGGTGTCCCGCGTGGCCCGAGTGTGCAAGAAC  
GACGTGGGAGGCTCCCCCGCGTGTGGAGAAGCAGTGGACGTCCTTCCTGAAGGCGCGGCTCAACTGCTCTGTA  
CCCCGAGACTCCCATTTCTACTTCAACGTGCTGCAGGCTGTACGGGCGTGGTCAGCCTCGGGGGCCGGCCCGTG  
GTCCTGGCCGTTTTTTCCACGCCCAGCAACAGCATCCCTGGCTCGGTGTCTGCGCCTTGACCTGACACAGGTG  
GCAGCTGTGTTTGAAGGCCGCTTCGAGAGCAGAAGTCCCCCGAGTCCATCTGGACGCCGGTGGCCGAGGATCAG  
GTCCCTCGAACCCCGGCCGGGTGCTGCGCAGCCCCGGGATGCAGTCAATGCCTCCAGCGCTTGCCGGATCAG  
ATCCTCAACTTTGTCAAGACCCACCCTCTGATGGACGAGGCGGTGCCCTCGCTGGGCCATGCGCCCTGGATCCTG  
CGGACCCTGATGAGGCACCAGCTGACTCGAGTGGCTGTGGACGTGGGAGCCGGCCCCCTGGGGCAACCAGACCGTT  
GTCTTCTGSGTTCTGAGGCGGGGACGGTCTCAAGTTCCTCGTCCGGCCCAATGCCAGCACCTCAGGGACGTCT  
GGGCTCAGTGTCTTCTGAGGAGTTTGAGACCTACCGGCCGACAGGTGTGGACGGCCCCGGCGGTGGCGAGACA  
GGGACGCGGCTGCTGAGCTTGGAGCTGGACGAGCTTGGGGGGCCTGCTGGCTGCCCTCCCCCGCTGCGTGTGC  
CGAGTGCCTGTGGTCTGCTGCCAGCAGTACTCGGGGTGTATGAAGAAGTGTATCGGCAGTCAGGACCCCTACTGC  
GGGTGGGCCCCCGACGGCTCCTGCATCTTCTCAGCCCGGGCACCAGAGCCGCTTTGAGCAGGACGTGTCCGGG  
GCCAGCACCTCAGGCTTAGGGGACTGCACAGGACTCCTGCGGGCCAGCCTCTCCGAGGACCGCGCGGGGCTGGTG  
TCGGTGAACCTGCTGGTAACGTGCTCGGTGGCGGCCCTCGTGGTGGGAGCCGTGGTGTCCGGCTTCAGCGTGGGC  
TGGTTCTGTGGGCTCCGTGAGCGGGGAGCTGGCCCGCGCAGAAGGACAAGGAGGCCATCTGGCGCACGGGGCG  
GGCAGGGCGGTGCTGAGCGTCAGCGCCTGGGCGAGCGCAGGGCGAGGTCCTCCGGGGCGGGGCGGAGCGGT  
GGCGGTGGCGCGGGGTTCCCCCGAGGCCCTGCTGGCGCCCTGATGCAGAACGGCTGGGCCAAGGCCACGCTG  
CTGCAGGGCGGGCCCCACGACCTGGACTCGGGGCTGCTGCCACGCCCCGAGCAGACGCCGCTGCCGCAAGCGC  
CTGCCACTCCGCACCGCACCCCCACGCCCTGGGCCCGCGCCTGGGACCACGGCCACCCCTGCTCCCGGCC  
TCCGCTTCATCTCCTCCTGCTGCTGGCGCCCGCGGCCCCCCGAGCAGCCCCCGCGCTGGGGAGCCGACC  
CCCCAGCGCGCCTCTATGCTGCCCGGCCCGCGCCTCCACCGCGACTTCCCGCTCACCCCCACGCGCAG  
CCGAGCCGCGCGGGTGGTGTCCGCGCCCCACGGGCCCTTGACCACGCTCAGCGCCGATGGCTCCCGCGG  
CCCTGGAGCCCGCCCCCGACGGGCAGCTGAGGAGGCCACTGGGCCCCACGCCCCCTCCGGCCGCCACCTGCGC  
CGCACCCACACGTTCAACAGCGGCGAGGCCCGGCTGGGGACCGCCACCGCGCTGCCACGCCCGGCCGGGCACA  
GACTTGGCCACCTCCTCCCTATGGGGGGGCGGACAGGACTGCGCCCCCGTGGCT**Tag**GCCGGGGGCCCCCCG  
ATGCTTTGGCAGTGCCAGCCACGGGAACCAGGAGCGAGAGACGGTGCCAGAAGCGCCGGGCGGGGCAACTCCG  
AGTGGGTGCTCAAGTCCCCCGGACCCACCGGAGTGGGGGGCCCCCTCCGCAAGGAAGCACAAACG  
CTCGCCCTCCCCCTACCGGGGCGCAGGACGCTGAGACGTTTTGGGGTGGGTGGCGGGAGGACTTTGCTATG  
GATTTGAGGTTGACCTTATGCGCGTAGGTTTTGGTTTTTTTTTTCAGTTTTGGTTTTCTTTTGGGTTTTCTAAC  
AATTGCACAACTCCGTTCTCGGGGTGGCGGCAGGCAGGGGAGGCTTGGACGCCGGTGGGGAATGGGGGGCCACAG  
CTGCAGACCTAAGCCCTCCCCCACCCTGGAAGGTCCCTCCCCAACCCAGGCCCTGGCGTGTGTGGGTGTGCG  
TGGCTGTGCGTGGCGTGTGTCAGGGGCGGGAGGTGGGCGTGTGTGCGTGCCAGCGAAGGCTGCTG  
TGGCGTGTGTGTCAGTGGGCGACGCTGAGGTTGTGTGCCACGAGCAGCATGCTGTTGGCCGCCAGCGGCC  
TGGGCGTTGGCTGAGCCGACGCTGGGGCTTCAGAAGGCCCGGGGGTCTCCGAGGTGCCGGTTAGGAGTTTGAAC  
CCCCCCTCTGACAGAGGAAGCGGGGACAATGCCGGGTTTCAGGCAGGAGACACGAGGAGGGCCTGCCCGGA  
AGTCACATCGGCAGCAGCTGTCTAAAGGGCTTGGGGGCTGGGGGGCGGCAAGGTGGGTGGGGCCCCCTCTGTAA  
ATACGGCCCCAGGGTGGTGAGAGAGTCCCATGCCACCCGCTCCCCTTGTGACCTCCCCCTATGACCTCCAGCTGA  
CCATGCATGCCAGCTGGCTGGTGGTCTTCTGCCCTCTTTGGAGTTTTGCCTCCCCAGCCCCCTCCCCATCAAT  
AAAACCTCTGTTTACAACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

## **FIGURE 16**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA80145

><subunit 1 of 1, 888 aa, 1 stop

><MW: 95285, pI: 8.89, NX(S/T): 8

MQTPRASPPRPALLLLLLLLGGAHGLFPEEPPPLSVAPRDYLNHYPVFVGSGPGRLTPAEGA  
DDLNIQRLVRVNRTLFIGDRDNLYRVELEPPTSTELRYQRKLTWRSNPSPDINVCRMKGKQEG  
ECRNFKVLLLRDESTLFVCGSNAFNPVCANYSIDTLQPVGDNISGMARCPYDPKHANVALF  
SDGMLFTATVTDFLAIDAVIYRSLGDRPTLRTVKHDSKWFKEPYFVHAVEWGSFVYFFFREI  
AMEFNYLEKVVVSRVARVCKNDVGGSPRVLEKQWTSFLKARLNCSVPGDSHFYFNVLQAVTG  
VVSLGGRPVVLAVFSTPSNSIPGSAVCAFDLTQVAAVFEGRFREQKSPESIWTVPVEDQVPR  
PRPGCCAAPGMQYNASSALPDDILNFVKTHPLMDEAVPSLGHAPWILRTLMRHQLTRVAVDV  
GAGPWGNQTVVFLGSEAGTVLKFLVRPNASTSGTSGLSVFLEEFETYRPPDRCGRPGGGETGQ  
RLLSLELDAASGGLLAAPRCVVRVPVARCQQYSGCMKNCIGSQDPYCGWAPDGCIFLSPG  
TRAAFEQDVSGASTSGLGDCTGLLRASLSEDRAGLVSVNLLVTSSVAAFVVGAVVSGFSVGW  
FVGLRERRELARRKDKEAILAHGAGEAVLSVSRLGERRAQGPGRGGGGGGGAGVPPEALLA  
PLMQNGWAKATLLQGGPHDLDSGLLPTPEQTPLPQKRLPTPHPHPHALGPRAWDHGHPLLP  
SASSLLLLAPARAPEQPPAPGEPTPDGRLYAARPGRASHGDFPLTPHASPDORRRVVSAPT  
PLDPASAADGLPRPWSPPPTGSLRRPLGPHAPPAATLRRTHTFNSGEARPGDRHRGCHARPG  
TDLAHLPLPYGGADRTAPPVP

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-25

#### **Transmembrane domains:**

amino acids 318-339, 598-617

#### **N-glycosylation sites.**

amino acids 74-78, 155-159, 167-171, 291-295, 386-390, 441-445,  
462-466

#### **Glycosaminoglycan attachment sites.**

amino acids 51-55, 573-577

#### **cAMP- and cGMP-dependent protein kinase phosphorylation site.**

amino acids 102-106

#### **N-myristoylation sites.**

amino acids 21-27, 50-56, 189-195, 333-339, 382-388, 448-454,  
490-496, 491-497, 508-514, 509-515, 531-537, 558-564, 569-575,  
574-580, 580-586, 610-616, 643-649, 663-669, 666-672, 667-673,  
668-674, 669-675, 670-676, 868-874, 879-885



## FIGURE 17

AGCAACTCAAGTTCATCATTGTCCTGAGAGAGAGGAGCAGCGCGGTTCTCGGCCGGGACAGC  
AGAACGCCAGGGGACCCTCACCTGGGCGCGCCGGGGCACGGGCTTTGATTGTCCTGGGGTCG  
CGGAGACCCGCGCGCCTGCCCTGCACGCCGGGCGGCAACCTTTGCAGTCGCGTTGGCTGCTG  
CGATCGGCCGGCGGGTCCCTGCCGAAGGCTCGGCTGCTTCTGTCCACCTCTTACACTTCTTC  
ATTTATCGGTGGATCATTTCGAGAGTCCGTCTTGTAATGTTTGGCACTTTGCTACTTTATT  
GCTTCTTTCTGGCGACAGTTCCAGCACTCGCCGAGACCGGCGGAGAAAGGCAGCTGAGCCCG  
GAGAAGAGCGAAATATGGGGACCCGGGCTAAAAGCAGACGTCGTCCTTCCCGCCCGCTATTT  
CTATATTCAGGCAGTGGATACATCAGGGAATAAATTACATCTTCTCCAGGCGAAAAGGTCT  
TCCAGGTGAAAGTCTCAGCACCAGAGGAGCAATTCACTAGAGTTGGAGTCCAGGTTTTAGAC  
CGAAAAGATGGGTCCTTCATAGTAAGATACAGAATGTATGCAAGCTACAAAAATCTGAAGGT  
GGAAATTAAATTCCAAGGGCAACATGTGGCCAAATCCCCATATATTTTAAAAGGGCCGGTTT  
ACCATGAGAACTGTGACTGTCTCTGCAAGATAGTGCAGCCTGGCTACGGGAGATGAACTGC  
CCTGAAACCATTGCTCAGATTCAGAGAGATCTGGCACATTTCCCTGCTGTGGATCCAGAAAA  
GATTGCAGTAGAAATCCCAAAAAGATTTGGACAGAGGCAGAGCCTATGTCACTACACCTTAA  
AGGATAACAAGGTTTATATCAAGACTCATGGTGAACATGTAGGTTTTAGAAATTTTCATGGAT  
GCCATACTACTTTCTTTGACTAGAAAGGTGAAGATGCCAGATGTGGAGCTCTTTGTTAATTT  
GGGAGACTGGCCTTTGGAAAAAAGAAATCCAATTCAAACATCCATCCGATCTTTTCCTGGT  
GTGGCTCCACAGATTCCAAGGATATCGTGATGCCTACGTACGATTTGACTGATTCTGTTCTG  
GAAACCATGGGCCGGGTAAAGTCTGGATATGATGTCCGTGCAAGCTAACACGGGTCCTCCCTG  
GGAAAGCAAAAATTCCTGCGTCTGGAGAGGGCGAGACAGCCGCAAAGAGAGACTCGAGC  
TGGTTAAACTCAGTAGAAAACACCCAGAACTCATAGACGCTGCTTTCACCAACTTTTTCTTC  
TTTAAACACGATGAAAACCTGTATGGTCCCATTTGTGAAACATATTTTCAATTTTTTGATTTCTT  
CAAGCATAAGTATCAAATAAATATCGATGGCACTGTAGCAGCTTATCGCCTGCCATATTTGC  
TAGTTGGTGACAGTGTTGTGCTGAAGCAGGATTCCATCTACTATGAACATTTTTTACAATGAG  
CTGCAGCCCTGGAAACACTACATTCCAGTTAAGAGCAACCTGAGCGATCTGCTAGAAAAACT  
TAAATGGGCGAAAGATCACGATGAAGAGGCCAAAAAGATAGCAAAAGCAGGACAAGAATTTG  
CAAGAAATAATCTCATGGGCGATGACATATTCTGTTATTATTTCAAACCTTTTCCAGGAATAT  
GCCAATTTACAAGTGAGTGAGCCCCAAATCCGAGAGGGCATGAAAAGGGTAGAACCACAGAC  
TGAGGACGACCTCTTCCCTTGTA CTGTCATAGGAAAAAGACCAAAGATGAACTCTGATATG  
CAAAATAACTTCTATTAGAATAATGGTGCTCTGAAGACTCTTCTTAACTAAAAAGAAGAATT  
TTTTTAAGTATTAATTCATGGACAATATAAAATCTGTGTGATTGTTTGCAGTATGAAGACA  
CATTTCTACTTATGCAGTATTCTCATGACTGTACTTTAAAGTACATTTTTTAGAATTTTATAA  
TAAAACCACCTTTATTTTAAAGGAAAAAA

## **FIGURE 18**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA84917

><subunit 1 of 1, 502 aa, 1 stop

><MW: 58043, pI: 7.94, NX(S/T): 2

MFGTLLLYCFFLATVPALAETGGERQLSPEKSEIWGPGLKADVVLARYFYIQAVDTSGNKF  
TSSPGEKVFQVKVSAPEEQFTRVGVQVLDRKDGSFIVRYRMYASYKNLKVVEIKFQGQHVAKS  
PYILKGPVYHENCDCPLQDSAAWLREMNCPETIAQIQRDLAHFPAVDPEKIAVEIPKRFGQR  
QSLCHYTLKDNKVYIKTHGEHVGFRIFMDAILLSLTRKVKMPDVELFVNLGDWPLEKKKSNS  
NIHPIFSWCGSTDSKDIVMPTYDLTDSVLETMGRVSLDMMSVQANTGPPWESKNSTAVWRGR  
DSRKERLELVKLSRKHPOLIDAAFTNFFFFKHDENLYGPIVKHISFFDFFKHKYQINIDGTV  
AAYRLPYLLVGDSVVLKQDSIYYEHFYNELQPWKHYIPVKSNLSDLLEKLKWAKDHDEEAKK  
IAKAGQEFARNNLMGDDIFCYYFKLFQEYANLQVSEPQIREGMKRVEPQTEDDLFPCTCHRK  
KTKDEL

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-17

#### **N-glycosylation sites.**

amino acids 302-306, 414-418

#### **cAMP- and cGMP-dependent protein kinase phosphorylation sites.**

amino acids 243-247, 495-499

#### **Tyrosine kinase phosphorylation site.**

amino acids 341-348

#### **N-myristoylation sites.**

amino acids 59-65, 118-124, 184-190, 258-264, 370-376, 439-445

#### **Endoplasmic reticulum targeting sequence.**

amino acids 499-504

## FIGURE 19

CCTGGAGCCGGAAGCGCGGCTGCAGCAGGGCGAGGCTCCAGGTGGGGTCGGTTCCGCATCCA  
GCCTAGCGTGTCCACG**ATG**CGGCTGGGCTCCGGGACTTTCGCTACCTGTTGCGTAGCGATCG  
AGGTGCTAGGGATCGCGGTCTTCCTTCGGGGATTCTTCCCGGCTCCCGTTCGTTCCCTCTGCC  
AGAGCGGAACACGGAGCGGAGCCCCAGCGCCCGAACCCCTCGGCTGGAGCCAGTTCTAACTG  
GACCACGCTGCCACCACCTCTCTTCAGTAAAGTTGTTATTGTTCTGATAGATGCCTTGAGAG  
ATGATTTTGTGTTTGGGTCAAAGGGTGTGAAATTTATGCCCTACACAACCTTACCTTGTGGAA  
AAAGGAGCATCTCACAGTTTTGTGGCTGAAGCAAAGCCACCTACAGTTACTATGCCTCGAAT  
CAAGGCATTGATGACGGGGAGCCTTCCTGGCTTTGTGACGTCATCAGGAACCTCAATTCTC  
CTGCACTGCTGGAAGACAGTGTGATAAGACAAGCAAAGCAGCTGGAAAAAGAATAGTCTTT  
TATGGAGATGAAACCTGGGTAAATTAATTCCTCCAAAGCATTTTGTGGAATATGATGGAACAAC  
CTCATTTTTCTGTGTCAGATTACACAGAGGTGGATAATAATGTCACGAGGCATTGGGATAAAG  
TATTA AAAAGAGAGGAGATTGGGACATATTAATCCTCCACTACCTGGGGCTGGACCACATTGGC  
CACATTTAGGGCCCAACAGCCCCCTGATTGGGCAGAAGCTGAGCGAGATGGACAGCGTGCT  
GATGAAGATCCACACCTCACTGCAGTCGAAGGAGAGAGAGACGCCTTTACCCAATTTGCTGG  
TTCTTTGTGGTGACCATGGCATGTCTGAAACAGGAAGTCACGGGGCCTCCTCCACCGAGGAG  
GTGAATACACCTCTGATTTTAATCAGTTCTGCGTTTGAAAGGAAACCCGGTGATATCCGACA  
TCCAAAGCACGTCCAAT**AG**ACGGATGTGGCTGCGACACTGGCGATAGCACTTGGCTTACCGA  
TTCCAAAAGACAGTGTAGGGAGCCTCCTATTCCAGTTGTGGAAGGAAGACCAATGAGAGAG  
CAGTTGAGATTTTTACATTTGAATACAGTGCAGCTTAGTAAACTGTTGCAAGAGAATGTGCC  
GTCATATGAAAAAGATCCTGGGTTTGAGCAGTTTAAATGTGAGAAAGATTGCATGGGAACT  
GGATCAGACTGTACTTGGAGGAAAAGCATTGAGAAAGTCTATTCAACCTGGGCTCCAAGGTT  
CTCAGGCAGTACCTGGATGCTCTGAAGACGCTGAGCTTGTCCCTGAGTGCACAAGTGGCCCA  
GTTCTCACCCCTGCTCCTGCTCAGCGTCCACAGGCAGTGCACAGAAAGGCTGAGCTGGAAGT  
CCCAGTGTATCTCCTGGGTTTTCTCTGCTCTTTTATTTGGTGATCCTGGTTCTTTTCGGCCG  
TTCACGTCATTGTGTGCACCTCAGCTGAAAGTTCGTGCTACTTCTGTGGCCTCTCGTGGCTG  
GCGGCAGGCTGCCTTTTCGTTTACCAGACTCTGGTTGAACACCTGGTGTGTGCCAAGTGTGG  
CAGTGCCCTGGACAGGGGGCCTCAGGGAAGGACGTGGAGCAGCCTTATCCCAGGCCTCTGGG  
TGTCCCGACACAGGTGTTACATCTGTGCTGTGAGGTGAGTGCCTCAGTTCTTGGAAGCT  
AGGTTCTGCGACTGTTACCAAGGTGATTGTAAAGAGCTGGCGGTCACAGAGGAACAAGCCC  
CCCAGCTGAGGGGGTGTGTGAATCGGACAGCCTCCCAGCAGAGGTGTGGGAGCTGCAGCTGA  
GGGAAGAAGAGACAATCGGCCTGGACACTCAGGAGGGTCAAAGGAGACTTGGTTCGCACCAC  
TCATCCTGCCACCCCCAGAATGCATCCTGCCTCATCAGGTCCAGATTTCTTTCCAAGGCGGA  
CGTTTTCTGTTGGAATTCTTAGTCCTTGGCCTCGGACACCTTCATTTCGTTAGCTGGGGAGTG  
GTGGTGAGGCAGTGAAGAAGAGGCGGATGGTCACTCAGATCCACAGAGCCCAGGATCAAG  
GGACCCACTGCAGTGGCAGCAGGACTGTTGGGCCCCCACCCTGCACAGCCCTCATC  
CCCTCTTGGCTTGAGCCGTCAGAGGCCCTGTGCTGAGTGTCTGACCGAGACACTCACAGCTT  
TGTCATCAGGGCACAGGCTTCCTCGGAGCCAGGATGATCTGTGCCACGCTTGACCTCGGGC  
CCATCTGGGCTCATGCTCTCTCTCTGCTATTGAATTAGTACCTAGCTGCACACAGTATGTA  
GTTACCAAAAAGAATAAACGGCAATAATTGAGAAAAAAA

## **FIGURE 20**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA84920

><subunit 1 of 1, 310 aa, 1 stop

><MW: 33875, pI: 7.08, NX(S/T): 2

MRLGSGTFATCCVAIEVLGIAVFLRGFFPAPVRSSARAEGAEPPAPEPSAGASSNWTTLPP  
PLFSKVIVLIDALRDDFVFGSKGVKFMPYTTYLVEKGASHSFVAEAKPPTVTMPRIKALMT  
GSLPGFVDVIRNLNNSPALLEDVIRQAKAAGKRIVFYGDETWVKLFPHFVEYDGTTTSFFVS  
DYTEVDNNVTRHLDKVLKRGDWDILILHYLGLDGHIGHISGPN SPLIGQKLSEMDSVLMKIHT  
SLQSKERETPLPNLLVLCGDHGMSETGSHGASSTEEVNTPLILISSAFERKPGDIRHPKHVQ

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-34

#### **Transmembrane domain:**

amino acids 58-76

#### **N-glycosylation sites.**

amino acids 56-60, 194-198

#### **N-myristoylation sites.**

amino acids 6-12, 52-58, 100-106, 125-131, 233-239, 270-276,  
275-281, 278-284

#### **Amidation site.**

amino acids 154-158

#### **Cell attachment sequence.**

amino acids 205-208

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[illegible][illegible]

## **FIGURE 22**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA86576

><subunit 1 of 1, 251 aa, 1 stop

><MW: 26935, pI: 7.42, NX(S/T): 2

MAMGVPRVILLCLFGAALCLTGSQALQCYSFEHTYFGPFDLRAMKLPSISCPHECFEAILSL  
DTGYRAPVTILVRKGCWTGPPAGQTQSNPDALPPDYSVVRGCTTDKCNAHLMTHDALPNLSQA  
PDPPTLSGAECYACIGVHQDDCAIGRSRRVQCHQDQTACFQSGRMTVGNFSVPVYIRTCHR  
PSCTTEGTTSPWTAIDLQGSCCEGYLCNRKSMTQPFTSASATTPPRALQVLALLLPVLLLVGLSA

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-19

#### **Transmembrane domain:**

amino acids 233-251

#### **N-glycosylation sites.**

amino acids 120-124, 174-178

#### **N-myristoylation sites.**

amino acids 15-21, 84-90

## FIGURE 23

CCCACGCGTCCGGGACAGATGAACTTAAAAGAGAAGCTTTAGCTGCCAAAGATTGGGAAAGG  
GAAAGGACAAAAAAGACCCCTGGGCTACACGGCGTAGGTGCAGGGTTTCCTACTGCTGTTCT  
TTTATGCTGGGAGCTGTGGCTGTAACCAACTAGGAAATAACGTATGCAGCAGCTATGGCTGT  
CAGAGAGTTGTGCTTCCCAAGACAAAGGCAAGTCCTGTTTCTTTTTCTTTTTTGGGGAGTGT  
CCTTGGCAGGTTCTGGGTTTGGACGTTATTCGGTGACTGAGGAAACAGAGAAAGGATCCTTT  
GTGGTCAATCTGGCAAAGGATCTGGGACTAGCAGAGGGGGAGCTGGCTGCAAGGGGAACCAAG  
GGTGGTTTCCGATGATAACAAACAATACCTGCTCCTGGATTACATACCGGGAATTTGCTCA  
CAAATGAGAACTGGACCGAGAGAAGCTGTGTGGCCCTAAAGAGCCCTGTATGCTGTATTTT  
CAAATTTTAATGGATGATCCCTTTCAGATTTACCGGGCTGAGCTGAGAGTCAGGGATATAAA  
TGATCACGCGCCAGTATTTTCAGGACAAAGAAACAGTCTTAAAAATATCAGAAAATACAGCTG  
AAGGGACAGCATTTAGACTAGAAAGAGCACAGGATCCAGATGGAGGACTTAACGGTATCCAA  
AACTACACGATCAGCCCCAACTCTTTTTTCCATATTAACATTAGTGGCGGTGATGAAGGCAT  
GATATATCCAGAGCTAGTGTGGACAAAGCACTGGATCGGGAGGAGCAGGGAGAGCTCAGCT  
TAACCCTCACAGCGCTGGATGGTGGGTCTCCATCCAGGTCTGGGACCTCTACTGTACGCATC  
GTTGTCTTGGACGTCAATGACAATGCCCCACAGTTTGCCAGGCTCTGTATGAGACCCAGGC  
TCCAGAAAACAGCCCCATTGGGTTCCCTTATTGTTAAGGTATGGGCAGAAGATGTAGACTCTG  
GAGTCAACGCGGAAGTATCCTATTCATTTTTTTGATGCCTCAGAAAATATTCGAACGACCTTT  
CAAATCAATCCTTTTTTCTGGGGAAATCTTCTCAGAGAATTGCTTGATTATGAGTTAGTAAA  
TTCTTACAAAATAAATATACAGGCAATGGACGGTGGAGGCCCTTCTGCAAGATGTAGGGTTT  
TAGTGGAAAGTATTGGACACCAATGACAATCCCCCTGAACTGATCGTATCATCATTTTTCCAAC  
TCTGTTGCTGAGAATTCTCCTGAGACGCCGCTGGCTGTTTTTAAGATTAATGACAGAGACTC  
TGGAGAAAATGGAAAGATGGTTTGCTACATTCAAGAGAATCTGCCATTCCCTACTAAAACCTT  
CTGTGGAGAATTTTTACATCCTAATTACAGAAGGCGCGCTGGACAGAGAGATCAGAGCCGAG  
TACAACATCACTATCACCGTCACTGACTTGGGGACACCCAGGCTGAAAACCGAGCACAAACAT  
AACGGTCTTGGTCTCCGACGTCAATGACAACGCCCCCGCCTTCACCCAAACCTCCTACACCC  
TGTTCTGTCGCGGAGAACAAACAGCCCCGCCCCGTCACATCGGCAGCGTCAGCGCCACAGACAGA  
GACTCGGGCACCAACGCCCAGGTCACTACTCGCTGCTGCCGCCCAAGACCCGACCTGCC  
CCTCGCCTCCCTGGTCTCCATCAACGCGGACAACGGCCACCTGTTTCGCCCTCAGGTGCTGG  
ACTACGAGGCCCTGCAGGCTTTCGAGTTCCGCGTGGGCGCCACAGACCGCGGCTCCCCCGCG  
CTGAGCAGAGAGGCGCTGGTGCGCGTGCTGGTGCTGGACGCCAACGACAACCTCGCCCTTCGT  
GCTGTACCCGCTGCAGAACGGCTCCGCGCCCTGCACCGAGCTGGTGCCCCGGGCGGCCGAGC  
CGGGCTACCTGGTGACCAAGGTGGTGGCGGTGGACGGCGACTCGGGCCAGAACGCCTGGCTG  
TCGTACCAGCTGCTCAAGGCCACGGAGCCCCGGGCTGTTCCGTGTGTGGGCGCACAAATGGGGA  
GGTGCGCACCGCCAGGCTGCTGAGCGAGCGCGACGCAGCCAAGCACAGGCTCGTGGTGCTTG  
TCAAGGACAATGGCGAGCCTCCTCGCTCGGCCACCGCCACGCTGCACTTGCTCCTGGTGGAC  
GGCTTCTCCAGCCCTACCTGCCTCTCCCGAGGCGGCCCGGCCAGGCCCAGGCCGAGGC  
CGACTTGCTCACCGTCTACCTGGTGGTGGCGTTGGCCTCGGTGTCTTCGCTCTTCCTCCTCT  
CGGTGCTCCTGTTCTGTGGCGGTGCGGCTGTGCAGGAGGAGCAGGGCGGCCTCGGTGGGTGCG  
TGCTCGGTGCCCGAGGGTCTTTTTCCAGGGCATCTGGTGGACGTGAGGGGCGCTGAGACCCT  
GTCCAGAGCTACCAGTATGAGGTGTGTCTGACGGGAGGCCCGGGACCAGTGAGTTCAAGT  
TCTTGAAACCAAGTTATTTTCGGATATTCAGGCACAGGGCCCTGGGAGGAAGGGTGAAGAAAAT  
TCCACCTTCCGAAATAGCTTTGGATTTAATATTCAGTAAAGTCTGTTTTTAGTTTCATATAC  
TTTTGGTGTGTTACATAGCCATGTTTCTATTAGTTTACTTTTAAATCTCAAATTTAAGTTAT  
TATGCAACTTCAAGCATTATTTTCAAGTAGTATACCCCTGTGGTTTTACAATGTTTCATCAT  
TTTTTTGCATTAATAACAACCTGGGTTTAAATTTAATGAGTATTTTTTTCTAAATGATAGTGT  
AAGGTTTTAATCTTTCCAACCTGCCCAAGGAATTAATTACTATTATATCTCATTACAGAAAT  
CTGAGGTTTTGATTCAATTTTCAGAGCTTGATCTCATGATTCTAATCACTTCTGTCTATAGTG  
TACTTGCTCTATTTAAGAAGGCATATCTACATTTCCAAACTCATTCTAACATTCTATATATT  
CGTGTTTGAAAACCATGTCATTTATTTCTACATCATGTATTTAAAAAGAAATATTTCTCTAC  
TACTATGCTCATGACAAAATGAAACAAAGCATATTGTGAGCAATACTGAACATCAATAATAC  
CCTTAGTTTATATACTTATTATTTTATCTTTAAGCATGCTACTTTTACTTGGCCAATATTTT  
CTTATGTTAACTTTTGCTGATGTATAAAACAGACTATGCCTTATAATTGAAATAAAATTATA  
ATCTGCCTGAAAATGAATAAAAATAAAACATTTTGAAATGTGAAAAAAAAAAAAAAAAAAAAA

## **FIGURE 24**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA87976

><subunit 1 of 1, 800 aa, 1 stop

><MW: 87621, pI: 4.77, NX(S/T): 7

MAVRELCFPRQRQVLFVFLFWGVSLAGSGFGRYSVTEETEKGSFVVNLAKDLGLAEGELAAR  
GTRVVSDDNKQYLLLDSTGNLLTNEKLDREKLCGPKEPCMLYFQILMDDPFQIYRAELRVR  
DINDHAPVFQDKETVLKISENTAEGTAFRLERAQDPDGGLNGIQNYTISPNSFFHINISGGD  
EGMIYPELVLDKALDREEQGELSLTLTALDGGSPSRSGTSTVRIVVLDVNDNAPQFAQALYE  
TQAPENSPIGFLIVKVWAEDVDSGVNAEVSYSFFDASENIRTTFQINPFSGEIFLRELLDYE  
LVNSYKINIQAMDGGGLSARCRVLVEVLDTNDNPPELIVSSFSNSVAENSPETPLAVFKIND  
RDSGENGMVCYIQENLPFLKPSVENFYILITEGALDREIRAENITITVTDLGTPLRKTE  
HNITVLVSDVNDNAPAFQTQTSYTLFVRENNSPALHIGSVSATDRDSGTNAQVTYSLLPPQDP  
HLPLASLV SINADNGHLFALRSLDYEALQAFEFVVGATDRGSPALSREALVRVLVLDANDNS  
PFVLYPLQNGSAPCTELVPRAAEPGYLVTKVAVDGDGQNAWLSYQLLKATEPGLFGVWAH  
NGEVRTARLLSERDAAKHRLVVLVKDNGEPPRSATATLHLLLVDFGFSQPYLPLPEAAPAQQAQ  
AEADLLTVYLVVALASVSSLFLLSVLLFVAVRLCRRSRAASVGRCSVPEGPFPGHLVDVRGA  
ETLSQSYQYEVCLTGPGTSEFKFLKPVISDIQAQGPGRKGEENSTFRNSFGFNIQ

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-26

#### **Transmembrane domain:**

amino acids 687-711

#### **N-glycosylation sites.**

amino acids 169-173, 181-185, 418-422, 436-440, 567-571, 788-792

#### **Glycosaminoglycan attachment site.**

amino acids 28-32

#### **Tyrosine kinase phosphorylation sites.**

amino acids 394-402, 578-585

#### **N-myristoylation sites.**

amino acids 22-28, 27-33, 53-59, 82-88, 162-168, 184-190,  
217-223, 324-330, 325-331, 471-477, 568-574, 759-765

#### **Amidation site.**

amino acids 781-785

#### **Aminoacyl-transfer RNA synthetases class-II signature 1.**

amino acids 117-138

#### **Cadherins extracellular repeated domain signature.**

amino acids 121-132, 230-241, 335-346, 439-450, 549-560





## **FIGURE 26**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA92234

><subunit 1 of 1, 507 aa, 1 stop

><MW: 56692, pI: 5.22, NX(S/T): 3

MDPKLGRMAASLLAVLLLLLLERGMFSSPSPPPALLEKVFQYIDLHQDEFVQTLKEWVAIESD  
SVQPVPRFRQELFRMMAVAADTLQRLGARVASVDMGPQQLPDGQSLPIPPVILAELGSDPTK  
GTVCFYGHLDVQPADRGDGLWLTDPYVLTEVDGKLYGRGATDNKGPVLAWINAVSAFRALEQD  
LPVNIKFIIIEGMEEAGSVALEELVEKEKDRFFSGVDYIVISDNLWISQRKPAITYGTRGNSY  
FMVEVKCRDQDFHSGTFGGILHEPMADLVALLGSLVDSSGHILVPGIYDEVVPLTEEEINTY  
KAIHLDLEEYRNSSRVEKFLFDTKEEILMHLWRYPSSLIHGIEGAFDEPGTKTVIPGRVIGK  
FSIRLVPHMNVSAVEKQVTRHLEDVFSKRNSSNKMVVSM TLGLHPWIANIDDTQYLAAKRAI  
RTVFGTEPDMIRDGSTIPIAKMFQEIVHKS VVLIPLGAVDDGEHSQNEKINRWNYIEGTKLF  
AAFFLEMAQLH

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-26

#### **Transmembrane domain:**

273-292

#### **N-glycosylation sites.**

amino acids 322-326, 382-386, 402-406

#### **cAMP- and cGMP-dependent protein kinase phosphorylation site.**

amino acids 400-404

#### **N-myristoylation sites.**

amino acids 89-95, 119-125, 162-168, 197-203, 242-248, 263-269,  
351-357

#### **Cell attachment sequence.**

amino acids 140-143

#### **ArgE / dapE / ACY1 / CPG:**

amino acids 156-167

## FIGURE 27

CTCGGCTGGATTTAAGGTTGCCGCTAGCCGCCTGGGAATTTAAGGGACCCACACTACCTTCC  
CGAAGTTGAAGGCAAGCGGTGATTGTTTGTAGACGGCGCTTTGT**CATG**GGACCTGTGCGGTT  
GGGAATATTGCTTTTCCTTTTTTTGGCCGTGCACGAGGCTTGGGCTGGGATGTTGAAGGAGG  
AGGACGATGACACAGAACGCTTGCCCAGCAAATGCGAAGTGTGTAAGCTGCTGAGCACAGAG  
CTACAGGCGGAACTGAGTCGCACCGGTGATCTCGAGAGGTGCTGGAGCTGGGGCAGGTGCT  
GGATACAGGCAAGAGGAAGAGACACGTGCCTTACAGCGTTTCAGAGACAAGGCTGGAAGAGG  
CCTTAGAGAATTTATGTGAGCGGATCCTGGACTATAGTGTTACGCTGAGCGCAAGGGCTCA  
CTGAGATATGCCAAGGGTCAGAGTCAGACCATGGCAACACTGAAAGGCCTAGTGCAGAAGGG  
GGTGAAGGTGGATCTGGGGATCCCTCTGGAGCTTTGGGATGAGCCCAGCGTGGAGGTCACAT  
ACCTCAAGAAGCAGTGTGAGACCATGTTGGAGGAGTTTGAAGACATTGTGGGAGACTGGTAC  
TTCCACCATCAGGAGCAGCCCCTACAAAATTTTCTCTGTGAAGGTCATGTGCTCCCAGCTGC  
TGAAACTGCATGTCTACAGGAACTTGGACTGGAAAGGAGATCACAGATGGGGGAAGAGAAAA  
CAGAAGGGGAGGAAGAGCAGGAGGAGGAGGAGGAAGAGGAGGAAGAGGAAGGGGGAGACAAG  
ATGACCAAGACAGGAAGCCACCCCAAACCTTGACCGAGAAGATCTT**TGA**CCCTTGCCTTTGAG  
CCCCCAGGAGGGGAAGGGATCATGGAGAGCCCTCTAAAGCCTGCACTCTCCCTGCTCCACAG  
CTTTCAGGGTGTGTTTATGAGTGA

CTGACTCCACCCAAGCTTGTAGCTGTTCTCTCCCATCTAACC  
TCAGGCAAGATCCTGGTGAAACAGCATGACATGGCTTCTGGGGTGGAGGGTGGGGGTGGAGG  
TCCTGCTCCTAGAGATGAACTCTATCCAGCCCCCTAATTGGCAGGTGTATGTGCTGACAGTA  
CTGAAAGCTTTCCTCTTTAACTGATCCCACCCCCACCCAAAAGTCAGCAGTGGCACTGGAGC  
TGTGGGCTTTGGGGAAGTCACTTAGCTCCTTAAGGTCTGTTTTTAGACCCTTCCAAGGAAGA  
GGCCAGAACGGACATTCTCTGCGATCTATATACATTGCCTGTATCCAGGAGGCTACACACCA  
GCAAACCGTGAAGGAGAATGGGACACTGGGTCATGGCCTGGAGTTGCTGATAATTTAGGTGG  
GATAGATACTTGGTCTACTTAAGCTCAATGTAACCCAGAGCCCACCATATAGTTTTATAGGT  
GCTCAACTTTCTATATCGCTATTAACTTTTTTCTTTTTTTCTA

## **FIGURE 28**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA92256

><subunit 1 of 1, 248 aa, 1 stop

><MW: 28310, pI: 4.63, NX(S/T): 0

MGPVRLGILLFLFLAVHEAWAGMLKEEDDDTERLPSKCEVCKLLSTELQAELSRTGRSREVL  
ELGQVLDTGKRKRHVPYSVSETRLEEALENLCERILDYSVHAERKGSRLRYAKGQSQTMATLK  
GLVQKGVKVDLGIPLELWDEPSVEVTYLKKQCETMLEEFEDIVGDWYFHHQEQLQNFLCEG  
HVLPAAEACQLQETWTGKEITDGEEKTEGEEEEEEEEEEEEEGGDKMTKTGSHPKLDREDL

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-21

#### **cAMP- and cGMP-dependent protein kinase phosphorylation site.**

amino acids 106-110

#### **N-myristoylation site.**

amino acids 115-121

#### **Amidation site.**

amino acids 70-74

[illegible]

AAGTACTTGTGTCCGGGTGGTGGACTGGATTAGCTGCGGAGCCCTGGAAGCTGCCTGTCCTT  
CTCCCTGTGCTTAACCAGAGGTGCCCATGGGTTGGACAATGAGGCTGGTCACAGCAGCACTG  
TTACTGGGTCTCATGATGGTGGTCACTGGAGACGAGGATGAGAACAGCCCGTGTGCCCATGA  
GGCCCTCTTGGACGAGGACACCCTCTTTTGCCAGGGCCTTGAAGTTTTCTACCCAGAGTTGG  
GGAACATTGGCTGCAAGGTTGTTCCCTGATTGTAACAACCTACAGACAGAAGATCACCTCCTGG  
ATGGAGCCGATAGTCAAGTTCCCGGGGGCCGTGGACGGCGCAACCTATATCCTGGTGTGGT  
GGATCCAGATGCCCCTAGCAGAGCAGAACCCAGACAGAGATTCTGGAGACATTGGCTGGTAA  
CAGATATCAAGGGCGCCGACCTGAAGAAAGGGAAGATTCAAGGGCCAGGAGTTATCAGCCTAC  
CAGGCTCCCTCCCCACCGGCACACAGTGGCTTCCATCGCTACCAGTTCTTTGTCTATCTTCA  
GGAAGGAAAAGTCATCTCTCTCCTTCCCAAGGAAAACAAAACCTCGAGGCTCTTGGAATGG  
ACAGATTTCTGAACCGCTTCCACCTGGGCGAACCTGAAGCAAGCACCCAGTTCATGACCCAG  
AACTACCAGGACTACCAACCCTCCAGGCTCCAGAGGAAGGGCCAGCGAGCCCAAGCACAA  
AACCAGGCAGAGATTAGCTGCCTGCTAGATAGCCGGCTTTGCCATCCGGGCATGTGGCCACAC  
TGCTCACCACCGACGATGTGGGTATGGAACCCCTCTGGATACAGAACCCCTTCTTTTCCAA  
ATTAAAAAAAAAAAAATCATCAA

## **FIGURE 30**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA92274

><subunit 1 of 1, 223 aa, 1 stop

><MW: 25402, pI: 8.14, NX(S/T): 1

MGWTMRLVTAALLLGLMMVVTGDEDENSPCAHEALLDEDTLFCQGLEVFYPELGNIGCKVVP  
DCNNYRQKITSWMEPIVKFPGAVDGATYILVMVDPDAPSRAEPRQRFWRHWLVTDIKGADLK  
KGKIQQQELSAYQAPSPPAHSGFHRYQFFVYLQEGKVISLLPKENKTRGSKMDRFLNRFHL  
GEPEASTQFMTQNYQDSPTLQAPRGRASEPKHKTRQR

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-22

#### **N-glycosylation site.**

amino acids 169-173

#### **Tyrosine kinase phosphorylation site.**

amino acids 59-68

#### **N-myristoylation sites.**

amino acids 54-60, 83-89, 130-136

#### **Phosphatidylethanolamine signature.**

amino acids 113-157

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

GTCGACCCACGCGTCCGAAGCTGCTGGAGCCACGATTCCAGTCCCCTGGACTGTAGATAAAGA  
 CCCTTTCTTGCCAGGTGCTGAGACAACCACACTATGAGAGGCACTCCAGGAGACGCTGATGG  
 TGGAGGAAGGGCCGTCTATCAATCAATCACTGTTGCTGTTATCACATGCAAGTATCCAGAGG  
 CTCTTGAGCAAGGCAGAGGGGATCCCATTTATTTGGGAATCCAGAATCCAGAAATGTGTTTG  
 TATTGTGAGAAGGTTGGAGAACAGCCCACATTGCAGCTAAAAGAGCAGAAGATCATGGATCT  
 GTATGGCCAACCCGAGCCCGTGAAACCCTTCCTTTTCTACCGTGCCAAGACTGGTAGGACCT  
 CCACCCTTGAGTCTGTGGCCTTCCCGGACTGGTTCATTGCCTCCTCCAAGAGAGACCAGCCC  
 ATCATTCTGACTTCAGAACTTGGGAAGTCATACAACACTGCCTTTGAATTAAATATAAATGA  
CTGAACTCAGCCTAGAGGTGGCAGCTTGGTCTTTGTCTTAAAGTTTCTGGTTCCCAATGTGT  
 TTTCTGTCTACATTTTCTTAGTGTCATTTTACGCTGGTGCTGAGACAGGAGCAAGGCTGCTG  
 TTATCATCTCATTTTATAATGAAGAAGAAGCAATTACTTCATAGCAACTGAAGAACAGGATG  
 TGGCCTCAGAAGCAGGAGAGCTGGGTGGTATAAGGCTGTCTCTCAAGCTGGTGCTGTGTAG  
 GCCACAAGGCATCTGCATGAGTGACTTTAAGACTCAAAGACCAAACACTGAGCTTTCTTCTA  
 GGGGTGGGTATGAAGATGCTTCAGAGCTCATGCGCGTTACCCACGATGGCATGACTAGCACA  
 GAGCTGATCTCTGTTTCTGTTTTGCTTTATTCCCTCTTGGGATGATATCATCCAGTCTTTAT  
 ATGTTGCCAATATACCTCATTTGTGTGTAATAGAACCTTCTTAGCATTAAAGACCTTGTAACA  
 AAAATAATTCTTGGGGTGGGTATGAAGATGCTTCAGAGCTCATGCGCGTTACCCACGATGGC  
 ATGACTAGCACAGAGCTGATCTCTGTTTCTGTTTTGCTTTATTCCCTCTTGGGATGATATCA  
 TCCAGTCTTTATATGTTGCCAATATACCTCATTTGTGTGTAATAGAACCTTCTTAGCATTAA  
 GACCTTGTAACAATAAATAAATTCTTGTGTTAAGTTAAATCATTTTTGTCTAATTGTAATGTG  
 TAATCTTAAAGTTAAATAAACTTTGTGTATTTATATAATAATAAAGCTAAAACCTGATATAAA  
 ATAAAGAAAGAGTAAACTG

## **FIGURE 32**

MRGTPGDADGGGRAVYQSITVAVITCKYPEALEQGRGDPIYLGIONPEMCLYCEKVGEQPTL  
QLKEQKIMDLYGQPEPVKPFIFYRAKTGRTSTLESVAFPDWFIASSKRDQPIILTSELGKSY  
NTAFELNIND

**Signal sequence:**

amino acids 1-17

**N-myristoylation site.**

amino acids 10-16

**Cell attachment sequence.**

amino acids 36-39

FIGURE 32



## **FIGURE 33**

GCGAGGCTGCACCAGCGCCTGGCACC**ATG**AGGACGCCTGGGCCTCTGCCCCGTGCTGCTGCTG  
CTCCTGGCGGGAGCCCCCGCCGCGCGGCCCACTCCCCGACCTGCTACTCCCGCATGCGGGC  
CCTGAGCCAGGAGATCACCCGCGACTTCAACCTCCTGCAGGTCTCGGAGCCCTCGGAGCCAT  
GTGTGAGATACCTGCCCAGGCTGTACCTGGACATACACAATTACTGTGTGCTGGACAAGCTG  
CGGGACTTTGTGGCCTCGCCCCCGTGTTGGAAAGTGGCCCAGGTAGATTCCCTTGAAGGACAA  
AGCACGGAAGCTGTACACCATCATGAACTCGTTCTGCAGGAGAGATTTGGTATTCCTGTTGG  
ATGACTGCAATGCCTTGAATAACCAATCCCAGTGACTACGGTCCTGCCAGATCGTCAGCGC  
**TAA**GGGAACTGAGACCAGAGAAAGAACCCAAGAGAACTAAAGTTATGTCAGCTACCCAGACT  
TAATGGGCCAGAGCCATGACCCTCACAGGTCTTGTGTTAGTTGTATCTGAAACTGTTATGTA  
TCTCTCTACCTTCTGGAAAACAGGGCTGGTATTCCTACCCAGGAACCTCCTTTGAGCATAGA  
GTTAGCAACCATGCTTCTCATTCCTTGACTCATGTCTTGCCAGGATGGTTAGATACACAGC  
ATGTTGATTTGGTCACTAAAAAGAAGAAAAGGACTAACAAGCTTCACTTTTATGAACAATA  
TTTTGAGAACATGCACAATAGTATGTTTTTATTACTGGTTTAAATGGAGTAATGGTACTTTTA  
TTCTTTCTTGATAGAAACCTGCTTACATTTAACCAAGCTTCTATTATGCCTTTTTCTAACAC  
AGACTTTCTTCACTGTCTTTCATTTAAAAAGAAATTAATGCTCTTAAGATATATATTTTACG  
TAGTGCTGACAGGACCCACTCTTTCATTGAAAGGTGATGAAAATCAAATAAAGAATCTCTTC  
ACATGGA

Figure 33

## **FIGURE 34**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA93011

><subunit 1 of 1, 136 aa, 1 stop

><MW: 15577, pI: 8.88, NX(S/T): 0

MRTPGPLPVLLLLLAGAPAARPTPPTCYSRMRALSQEITRDFNLLQVSEPSEPCVRYLPRLY  
LDIHNYCVLDKLRDFVASPPCWKVAQVDSLKDKARKLYTIMNSFCRRDLVFLDDCNALEY  
IPVTTVLPDRQR

**Important features of the protein:**

**Signal peptide:**

amino acids 1-19

**Tyrosine kinase phosphorylation site.**

amino acids 60-69

**N-myristoylation site.**

amino acids 16-22

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

## FIGURE 35

GTCTCCGCGTCACAGGAACCTTCAGCACCCACAGGGCGGACAGCGCTCCCCTCTACCTGGAGA  
CTTGACTCCCGCGCGCCCCAACCTGCTTATCCCTTGACCGTCGAGTGTGAGAGATCCTGCA  
GCCGCCAGTCCCGGGCCCTCTCCCGCCCCACACCCACCCTCCTGGCTCTTCCTGTTTTTAC  
TCCTCCTTTTTTCATTCATAACAAAAGCTACAGCTCCAGGAGCCCAGCGCCGGGCTGTGACCCA  
AGCCGAGCGTGGAAGAATGGGGTTCCTCGGGACCGGCACTTGGAATTCTGGTGTTAGTGCTCC  
CGATTCAAGCTTTCCCCAAACCTGGAGGAAGCCAAGACAAATCTCTACATAATAGAGAATTA  
AGTGCAGAAAGACCTTTGAATGAACAGATTGCTGAAGCAGAAGAAGACAAGATTAAAAAAC  
ATATCCTCCAGAAAACAAGCCAGGTCAGAGCAACTATTCTTTTGTGATAACTTGAACCTGC  
TAAAGGCAATAACAGAAAAGGAAAAAATTGAGAAAGAAAGACAATCTATAAGAAGCTCCCCA  
CTTGATAATAAGTTGAATGTGGAAGATGTTGATTCAACCAAGAATCGAAAACCTGATCGATGA  
TTATGACTCTACTAAGAGTGGATTGGATCATAAATTTCAAGATGATCCAGATGGTCTTCATC  
AACTAGACGGGACTCCTTTAACCGCTGAAGACATTGTCCATAAAATCGCTGCCAGGATTTAT  
GAAGAAAATGACAGAGCCGTGTTTGACAAGATTGTTTCTAAACTACTTAATCTCGGCCTTAT  
CACAGAAAGCCAAGCACATACACTGGAAGATGAAGTAGCAGAGGTTTTACAAAATTAATCT  
CAAAGGAAGCCAACAATTATGAGGAGGATCCCAATAAGCCCACAAGCTGGACTGAGAATCAG  
GCTGGAAAAATACCAGAGAAAGTGAATCCAATGGCAGCAATTCAAGATGGTCTTGCTAAGGG  
AGAAAACGATGAAACAGTATCTAACACATTAACCTTGACAAATGGCTTGGAAGGAGAACTA  
AAACCTACAGTGAAGACAACCTTTGAGGAACTCCAATATTTCCCAAATTTCTATGCGCTACTG  
AAAAGTATTGATTGAGAAAAAGAAGCAAAAGAGAAAGAAACACTGATTACTATCATGAAAAC  
ACTGATTGACTTTGTGAAGATGATGGTGAAATATGGAACAATATCTCCAGAAGAAGGTGTTT  
CCTACCTTGAAAACCTTGGATGAAATGATTGCTCTTCAGACCAAAAACAAGCTAGAAAAAAT  
GCTACTGACAATATAAGCAAGCTTTTCCCAGCACCATCAGAGAAGAGTCATGAAGAAACAGA  
CAGTACCAAGGAAGAAGCAGCTAAGATGGAAAAGGAATATGGAAGCTTGAAGGATTCCACAA  
AAGATGATAACTCCAACCCAGGAGGAAAGACAGATGAACCCAAAGGAAAAACAGAAGCCTAT  
TTGGAAGCCATCAGAAAAAATATTGAATGGTTGAAGAAACATGACAAAAAGGGAAATAAAGA  
AGATTATGACCTTTCAAAGATGAGAGACTTCATCAATAAACAAGCTGATGCTTATGTGGAGA  
AAGGCATCCTTGACAAGGAAGAAGCCGAGGCCATCAAGCGCATTTATAGCAGCCTGTAAAAA  
TGGCAAAAGATCCAGGAGTCTTTCAACTGTTTCAGAAAACATAATATAGCTTAAAACACTTC  
TAATTCTGTGATTAAATTTTTTTGACCCAAGGGTTATTAGAAAGTGCTGAATTTACAGTAGT  
TAACCTTTTACAAGTGGTTAAAACATAGCTTTCTTCCCGTAAAACTATCTGAAAGTAAAGT  
TGTATGTAAGCTGAAAAAAAAAAAAAAAAAAAAA

## **FIGURE 36**

MGFLGTGTWILVVLVLP IQAFPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKIKKTYPPEN  
KPGQSNYSFVDNLNLLKAITEKEKIEKERQSIRSSPLDNKLNVEDVDSTKNRKLIDDYDSTK  
SGLDHKFQDDPDGLHQLDGTPLTAEDIVHKIAARIYEENDRAVFDKIVSKLLNLGLITESQA  
HTLEDEVAEVLQKLISKEANNYEEDPNKPTSWTENQAGK IPEKVTPMAAIQDGLAKGENDET  
VSNTLTLTNGLERRTKTYSEDNFEELQYFPNFYALLKSIDSEKEAKEKETLITIMKTLIDFV  
KMMVKYGTISPEEGVSYLENLDEMIALQTKNKLEKNATDNISKLF PAPSEKSHEETDSTKEE  
AAKMEKEYGSLKDSTKDDNSNPGGKTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLS  
KMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL

### **N-glycosylation sites:**

amino acids 68-71, 346-349, 350-353

### **Casein kinase II phosphorylation site:**

amino acids 70-73, 82-85, 97-100, 125-128, 147-150, 188-191, 217-  
220, 265-268, 289-292, 305-308, 320-323, 326-329, 362-365, 368-  
341, 369-372, 382-385, 386-389, 387-390

### **N-myristoylation sites:**

amino acids 143-148, 239-244

Figure 1 consists of 12 histograms arranged in a single column, labeled  $x_0$  through  $x_{11}$  on the right. Each histogram shows the frequency of values for  $x_k$  ranging from 0 to 10. The y-axis for each histogram is labeled 'Frequency' and ranges from 0 to 10. The distributions are roughly bell-shaped and centered around 5, with the peak frequency increasing from 10 at  $k=0$  to 12 at  $k=11$ .

## **FIGURE 38**

MALPPGPAALRHTLLLLPALLSSGWGELEPQIDGQTWAERALRENERHAFTCRVAGGPGT  
PR  
LAWYLDGQLQEASTSRLLSVGGEAFSGGTSTFTVTAHRAQHELNCSLQDPRSGRSANASVIL  
NVQFKPEIAQVGAKYQEAQGPGLLVVLFALVRANPPANVTWIDQDGPVTVNTSDFLVLDAQN  
YPWLTNHTVQLQLRSLAHNLSVVATNDVGVTASALPAPGPSRHPSLISSDSNNLKLNNVRLP  
RENMSLPSNLQLNDLTPDSRAVKPADRQMAQNNSRPELLDPEPGGLLTSQGFIRLPVLGYIY  
RVSSVSSDEIWL

### **N-glycosylation sites:**

amino acids 106-110, 119-123, 162-166, 175-179, 192-196, 205-209,  
251-255, 280-284

### **Glycosaminoglycan attachment site:**

amino acids 23-27

### **Casein kinase II phosphorylation sites:**

amino acids 36-40, 108-112, 164-168, 282-286, 316-320

### **N-myristoylation sites:**

amino acids 34-40, 89-95, 215-221, 292-298, 293-299

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## **FIGURE 39**

CGGGGACGGAAGCGGCCCTGGGCCCAGGGGCTGGAGCCGGGCCGGGGCGATGTGGAGCGC  
GGGCCGCGGCGGGGCTGCCTGGCCGGTGCTGTTGGGGCTGCTGCTGGCGCTGTTAGTGCCGG  
GCGGTGGTGCCGCCAAGACCGGTGCGGAGCTCGTGACCTGCGGGTCGGTGCTGAAGCTGCTC  
AATACGCACCACCGCGTGCGGCTGCACTCGCACGACATCAAATACGGATCCGGCAGCGGCCA  
GCAATCGGTGACCGGCGTAGAGGCGTCGGACGACGCCAATAGCTACTGGCGGATCCGCGGCG  
GCTCGGAGGGCGGGTGCCCGCGCGGGTCCCCGGTGCGCTGCGGGCAGGCGGTGAGGCTCACG  
CATGTGCTTACGGGCAAGAACCTGCACACGCACCACTTCCCGTCGCCGCTGTCCAACAACCA  
GGAGGTGAGTGCCTTTGGGGAAGACGGCGAGGGCGACGACCTGGACCTATGGACAGTGCCT  
GCTCTGGACAGCACTGGGAGCGTGAGGCTGCTGTGCGCTTCCAGCATGTGGGCACCTCTGTG  
TTCCTGTCAGTCACGGGTGAGCAGTATGGAAGCCCCATCCGTGGGCAGCATGAGGTCCACGG  
CATGCCCAGTGCCAACACGCACAATACGTGGAAGGCCATGGAAGGCATCTTCATCAAGCCTA  
GTGTGGAGCCCTCTGCAGGTCACGATGAACTCTGAGTGTGTGGATGGATGGGTGGATGGAGG  
GTGGCAGGTGGGGCGTCTGCAGGGCCACTCTTGGCAGAGACTTTGGGTTTGTAGGGGTCCTC  
AAGTGCCTTTGTGATTAAAGAATGTTGGTCTATGAAA

CGGGGACGGAAGCGGCCCTGGGCCCAGGGGCTGGAGCCGGGCCGGGGCG

## **FIGURE 40**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA96857

><subunit 1 of 1, 221 aa, 1 stop

><MW: 23598, pI: 6.96, NX(S/T): 0

MWSAGRGGAAWPVLLGLLLALLVPGGGAAKTGAELVTCGSVLKLLNTHHRVRLHSHDIKYGS  
GSGQQSVTGVEASDDANSYWRIRGGSEGGCPRGSPVRCGQAVRLTHVLTGKNLHTHHFPSPL  
SNNQEVSAFGEDGEGDDLDLWTVRCSGQHWEREA AVRFQHVGT SVFLSVTGEQYGSPIRGQH  
EVHGMPSANTHNTWKAMEGIFIKPSVEPSAGHDEL

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-28

#### **Glycosaminoglycan attachment site.**

amino acids 62-66

#### **N-myristoylation sites.**

amino acids 16-22, 25-31, 27-33, 61-67, 71-77, 86-92, 87-93,  
91-97, 190-196

#### **Endoplasmic reticulum targeting sequence.**

amino acids 218-223

FIGURE 40





## **FIGURE 42**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA96867

><subunit 1 of 1, 194 aa, 1 stop

><MW: 21431, pI: 8.57, NX(S/T): 0

MSALWLLLGLLALMDLSESSNWGCYGNIQSLDTPGASCGIGRRHGLNYCGVRASERLAEIDM  
PYLLKYQPMMQTIGQKYCMDPAVIAGVLSRKSPGDKILVNMGDRTSMVQDPGSQAPTSWISE  
SQVSQTTEVLTTRIKEIQRRFPTWTPDQYLRGGLCAYSGGAGYVRSSQDLSCDFCNDVLARA  
KYLKRHGF

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-19

#### **N-myristoylation sites.**

amino acids 23-29, 26-32, 35-41, 45-51, 50-56, 76-82, 156-162

#### **Amidation site.**

amino acids 40-44

protein = 194 aa

## **FIGURE 43**

TTGAAAATCTACTCTATCAGCTGCTGTGGTTGCCACCATTCTCAGGACCCTCGCC**ATG**AAAG  
CCCTTATGCTGCTCACCTGTCTGTTCTGCTCTGCTGGGTCTCAGCTGACATTGCTGTGCTCAC  
TCCTGCTACAAGGTCCCTGTGCTGGGCTGTGTGGACCGGCAGTCCTGCCGCCTGGAGCCAGG  
ACAGCAATGCCTGACAACACATGCATACCTTGGTAAGATGTGGGTTTTCTCCAATCTGCGCT  
GTGGCACACCAGAAGAGCCCTGTCAGGAGGCCTTCAACCAAACCAACCGCAAGCTGGGTCTG  
ACATATAACACCACCTGCTGCAACAAGGACAACCTGCAACAGCGCAGGACCCCGGCCCACTCC  
AGCCCTGGGCCTTGTCTTCCTTACCTCCTTGGCTGGCCTTGGCCTCTGGCTGCTGCAC**TGAG**  
ACTCATTCCATTGGCTGCCCCCTCCTCCCACCTGCCTTGGCCTGAGCCTCTCTCCCTGTGTCT  
CTGTATCCCCTGGCTTTACAGAATCGTCTCTCCCTAGCTCCCATTTCTTTAATTAAACACTG  
TTCCGAGTGGTCTCCTCATCCATCCTTCCCACCTCACACCCTTCACTCTCCTTTTTCTGGGT  
CCCTTCCCCTTCCCTTCCAGGACCTCCATTGGCTCCTAGAAGGGCTCCCCACTTTGCTTCCT  
ATACTCTGCTGTCCCCTACTTGAGGAGGGATTGGGATCTGGGCCTGAAATGGGGCTTCTGTG  
TTGTCCCCAGTGAAGGCTCCCACAAGGACCTGATGACCTCACTGTACAGAGCTGACTCCCCA  
AAGCCAGGCTCCCATATGTACCCCATCCCCATACTCACCTCTTTCCATTTTGAGTAATAAA  
TGTCTGAGTCTGGAAAAAAAAAAAAAAAAAAAA

TTGAAAATCTACTCTATCAGCTGCTGTGGTTGCCACCATTCTCAGGACCCTCGCC**ATG**AAAG

><MW: 13821, pI: 8.60, NX(S/T): 2

amino acids 85-98

[illegible]

## FIGURE 45

ACGGGCCG CAGCGGCAGTGACGTAGGGTTGGCGCACGGATCCGTTGCGGCTGCAGCTCTGCA  
GTCGGGCCGTTCTTCGCCGCCGCCAGGGGTAGCGGTGTAGCTGCGCAGCGTCGCGCGCGCT  
ACCGCACCCAGGTTTCGGCCCGTAGGCGTCTGGCAGCCCCGGCGCCATCTTCATCGAGCGCCAT  
GGCCGCAGCCTGCGGGCCGGGAGCGGCCGGGTACTGCTTGCTCCTCGGCTTGCATTTGTTTC  
TGCTGACCGCGGGCCCTGCCCTGGGCTGGAACGACCCTGACAGAATGTTGCTGCGGGATGTA  
AAAGCTCTTACCCTCCACTATGACCGCTATACCACCTCCCGCAGGCTGGATCCCATCCCACA  
GTTGAAATGTGTTGGAGGCACAGCTGGTTGTGATTCTTATACCCCAAAGTCATACAGTGTC  
AGAACAAAGGCTGGGATGGGTATGATGTACAGTGGGAATGTAAGACGGACTTAGATATTGCA  
TACAAATTTGAAAACTGTGGTGAGCTGTGAAGGCTATGAGTCCTCTGAAGACCAGTATGT  
ACTAAGAGGTTCTTGTGGCTTGGAGTATAATTTAGATTATACAGAACTTGGCCTGCAGAAAC  
TGAAGGAGTCTGGAAAGCAGCACGGCTTTGCCTCTTTCTCTGATTATTATTATAAGTGGTCC  
TCGGCGGATTCCTGTAACATGAGTGGATTGATTACCATCGTGGTACTCCTTGGGATCGCCTT  
TGTAAGTCTATAAGCTGTTCTGAGTGACGGGCAGTATTCTCCTCCACCGTACTCTGAGTATC  
CTCCATTTTCCCACCGTTACCAGAGATTCACCAACTCAGCAGGACCTCCTCCCCCAGGCTTT  
AAGTCTGAGTTCACAGGACCACAGAATACTGGCCATGGTGCAACTTCTGGTTTTTGGCAGTGC  
TTTTACAGGACAACAAGGATATGAAAATTCAGGACCAGGGTTCTGGACAGGCTTGGGAACTG  
GTGGAATACTAGGATATTTGTTTGGCAGCAATAGAGCGGCAACACCCTTCTCAGACTCGTGG  
TACTACCCGTCCTATCCTCCCTCCTACCCTGGCACGTGGAATAGGGCTTACTCACCCCTTCA  
TGGAGGCTCGGGCAGCTATTCGGTATGTTCAAACCTCAGACACGAAAACCAGAACTGCATCAG  
GATATGGTGGTACCAGGAGACGATTAAAGTAGAAAGTTGGAGTCAAACACTGGATGCAGAAAT  
TTTGGATTTTTTCATCACTTTCTCTTTAGAAAAAAAGTACTACCTGTTAACAATTGGGAAAAG  
GGGATATTCAAAGTTCTGTGGTGTTATGTCCAGTGTAGCTTTTTGTATTCTATTATTTGAG  
GCTAAAAGTTGATGTGTGACAAAATACTTATGTGTTGTATGTCAGTGTAACATGCAGATGTA  
TATTGCAGTTTTTGAAGTGATCACTTACTGTGGAATGCTAAAAATACATTAATTTCTAAAC  
CTGTGATGCCCTAAGAAGCATTAGAATGAAGGTGTTGTACTAATAGAACTAAGTACAGAA  
AATTTAGTTTTAGGTGGTTGTAGCTGATGAGTTATTACCTCATAGAGACTATAATATTCTA  
TTTGGTATTATATTATTTGATGTTTGCTGTTCTTCAAACATTTAAATCAAGCTTTGGACTAA  
TTATGCTAATTTGTGAGTTCTGATCACTTTTGAGCTCTGAAGCTTTGAATCATTCAGTGGTG  
GAGATGGCCTTCTGGTAACTGAATATTACCTTCTGTAGGAAAAGGTGGAAAATAAGCATCTA  
GAAGGTGTTGTGAATGACTCTGTGCTGGCAAAAATGCTTGAAACCTCTATATTTCTTTTCGT  
TCATAAGAGGTAAAGGTCAAATTTTTCAACAAAAGTCTTTTAATAACAAAAGCATGCAGTTC  
TCTGTGAAATCTCAAATATTGTTGTAATAGTCTGTTTCAATCTTAAAAAGAATCA

## **FIGURE 46**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA96889

><subunit 1 of 1, 339 aa, 1 stop

><MW: 36975, pI: 7.85, NX(S/T): 1

MAAACGPGAAGYCLLLGLHLFLLTAGPALGWNDPDRMLLRDVKALTLHYDRYTTSRRLDPIP  
QLKCVGGTAGCDSYTPKVIQCQNKGWDGYDVQWECKTDLDIAYKFGKTVVSCEGYESSEDQY  
VLRGSCGLEYNLDYTELGQLKESGKQHGFAFSDYKKWSSADSCNMSGITIVVLLGIA  
FVVYKFLFLSDGQYSPPPYSEYPPFSHRYQRFTNSAGPPPPGFKSEFTGPQNTGHGATSGFGS  
AFTGQQGYENSGPGFWTGLGTGGILGYLFGSNRAATPFSDSWYYPSYPPSYPGTWNRAYSPL  
HGGSGSYSVCSNSDTKTRTASGYGGTRRR

### **Signal peptide:**

amino acids 1-30

### **Transmembrane domain:**

amino acids 171-190

### **N-glycosylation site.**

amino acids 172-176

### **Glycosaminoglycan attachment sites.**

amino acids 244-248, 259-263, 331-335

### **Tyrosine kinase phosphorylation site.**

amino acids 98-106

### **N-myristoylation sites.**

amino acids 68-74, 69-75, 131-137, 241-247, 247-253, 266-272,  
270-276, 278-284, 312-318